

## **Hand Gesture Interaction with Touch Surface**

### **Field of the Invention**

[01] This invention relates generally to touch sensitive surfaces, and more particularly to using touch surfaces to recognize and act upon hand gestures made by touching the surface.

### **Background of the Invention**

[02] Recent advances in sensing technology have enabled increased expressiveness of freehand touch input, see Ringel et al., “Barehands: Implement-free interaction with a wall-mounted display,” Proc CHI 2001, pp. 367-368, 2001, and Rekimoto “SmartSkin: an infrastructure for freehand manipulation on interactive surfaces,” Proc CHI 2002, pp. 113-120, 2002.

[03] A large touch sensitive surface presents some new issues that are not present with traditional touch sensitive devices. Any touch system is limited by its sensing resolution. For a large surface, the resolution can be considerably lower than that with traditional touch devices. When each one of multiple users can simultaneously generate multiple touches, it becomes difficult to determine a context of the touches. This problem has been addressed, in part, for single inputs, such as for mouse-based and pen-based stroke gestures, see André et al., “Paper-less editing and proofreading of electronic documents,” Proc. EuroTeX, 1999, Guimbretiere et al., “Fluid Interaction with high-resolution wall-size displays. Proc. UIST 2001, pp. 21-30, 2001, Hong et al., “SATIN: A toolkit for informal ink-based applications,” Proc. UIST 2000, pp. 63-72, 2001, Long et al.,

“Implications for a gesture design tool,” Proc. CHI 1999, pp. 40-47, 1999, and Moran et al., “Pen-based interaction techniques for organizing material on an electronic whiteboard,” Proc. UIST 1997, pp. 45-54, 1992.

[04] The problem becomes more complicated for hand gestures, which are inherently imprecise and inconsistent. A particular hand gesture for a particular user can vary over time. This is partially due to the many degrees of freedom in the hand. The number of individual hand poses is very large. Also, it is physically demanding to maintain the same hand pose over a long period of time.

[05] Machine learning and tracking within vision-based systems have been used to disambiguate hand poses. However, most of those systems require discrete static hand poses or gestures, and fail to deal with highly dynamic hand gestures, Cutler et al., “Two-handed direct manipulation on the responsive workbench,” Proc I3D 1997, pp. 107-114, 1997, Koike et al., “Integrating paper and digital information on EnhancedDesk,” ACM Transactions on Computer-Human Interaction, 8 (4), pp. 307-322, 2001, Krueger et al., “VIDEOPLACE - An artificial reality, Proc CHI 1985, pp. 35-40, 1985, Oka et al., “Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems,” Proc FG 2002, pp. 429-434, 2002, Pavlovic et al., “Visual interpretation of hand gestures for human-computer interaction: A review,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 19 (7). pp. 677-695, 1997, and Ringel et al., “Barehands: Implement-free interaction with a wall-mounted display,” Proc CHI 2001, pp. 367-368, 2001. Generally, camera-based systems are difficult and expensive to implement, require extensive calibration, and are typically confined to controlled settings.

[06] Another problem with an interactive touch surface that also displays images is occlusion. This problem has been addressed for single point touch screen interaction, Sears et al., "High precision touchscreens: design strategies and comparisons with a mouse," International Journal of Man-Machine Studies, 34 (4). pp. 593-613, 1991 and Albinsson et al., "High precision touch screen interaction," Proc CHI 2003, pp. 105-112, 2003. Pointers have been used to interact with wall-based display surfaces, Myers et al., "Interacting at a distance: Measuring the performance of laser pointers and other devices," Proc. CHI 2002, pp. 33-40, 2002.

[07] It is desired to provide a gesture input system for a touch sensitive surface that can recognize multiple simultaneous touches by multiple users.

### **Summary of the Invention**

[08] It is an object of the invention to recognize different hand gestures made by touching a touch sensitive surface.

[09] It is desired to recognize gestures made by multiple simultaneous touches.

[010] It is desired to recognize gestures made by multiple users touching a surface simultaneously.

[011] A method according to the invention recognizes hand gestures. An intensity of a signal at touch sensitive pads of a touch sensitive surface is measured. The number of regions of contiguous pads touched simultaneously is determined from the intensities of the signals. An area of each region is determined. Then, a particular gesture is selected according to the number of regions touched and the area of each region.

## **Brief Description of the Drawings**

[012] Figure 1 is a block diagram of a touch surface for recognizing hand gestures according to the invention;

[013] Figure 2A is a block diagram of a gesture classification process according to the invention;

[014] Figure 2B is a flow diagram of a process for performing gesture modes;

[015] Figure 3 is a block diagram of a touch surface and a displayed bounding box;

[016] Figure 4 is a block diagram of a touch surface and a displayed bounding circle; and

[017] Figures 5-9 are examples hand gestures recognized by the system according to the invention.

## **Detailed Description of the Preferred Embodiment**

[018] The invention uses a touch surface to detect hand gestures, and to perform computer operations according to the gestures. We prefer to use a touch surface that is capable of recognizing simultaneously multiple points of touch from multiple users, see Dietz et al., “DiamondTouch: A multi-user touch technology,” Proc. User Interface Software and Technology (UIST) 2001, pp. 219-226, 2001, and U.S. Patent No. 6,498,590 “Multi-user touch surface,” issued to Dietz et al., on December 24, 2002, incorporated herein by reference. This touch surface can be made arbitrarily large, e.g.,

the size of a tabletop. In addition, it is possible to project computer generated images on the surface during operation.

[019] By gestures, we mean moving hands or fingers on or across the touch surface. The gestures can be made by one or more fingers, by closed fists, or open palms, or combinations thereof. The gestures can be performed by one user or multiple simultaneous users. It should be understood that other gestures than the example gestures described herein can be recognized.

[020] The general operating framework for the touch surface is described in U.S. Patent Application 10/053,652 “Circular Graphical User Interfaces” filed by Vernier et al., on January 18 2002, incorporated herein by reference. Single finger touches can be reserved for traditional mouse-like operations, e.g., point and click, select, drag, and drop, as described in the Vernier application.

[021] Figure 1 is used to describe the details of operation of the invention. A touch surface 100 includes  $m$  rows 101 and  $n$  columns 102 of touch sensitive pads 105, shown enlarged for clarity. The pads are diamond-shaped to facilitate the interconnections. Each pad is in the form of an antenna that couples capacitively to a user when touched, see Dietz above for details. The signal intensity of a single pad can be measured.

[022] Signal intensities 103 of the coupling can be read independently for each column along the  $x$ -axis, and for each row along the  $y$ -axis. Touching more pads in a particular row or column increases the signal intensity for that row or column. That is, the measured signal is proportional to the number of pads touched. It is observed that the signal intensity is generally greater in the middle part of a finger touch because of a better

coupling. Interestingly, the coupling also improves by applying more pressure, i.e., the intensity of the signal is coarsely related to touching pressure.

**[023]** The rows and columns of antennas are read along the  $x$ - and  $y$ -axis at a fixed rate, e.g., 30 frames/second, and each reading is presented to the software for analysis as a single vector of intensity values  $(x_0, x_1, \dots, x_m, y_0, y_1, \dots, y_n)$ , for each time step. The intensity values are thresholded to discard low intensity signals and noise.

**[024]** In Figure 1, the bold line segments indicate the corresponding  $x$  and  $y$  coordinates of the columns and rows, respectively that have intensities 104 corresponding to touching. In the example shown, two fingers 111-112 touch the surface. The signal intensities of contiguously touched rows of antennas are summed, as are signals of contiguously touched columns. This enables one to determine the number of touches, and an approximate area of each touch. It should be noted that in the prior art, the primary feedback data are  $x$  and  $y$  coordinates, i.e., a location of a zero dimensional point. In contrast, the primary feedback is a size of an area of a region touched. In addition, a location can be determined for each region, e.g., the center of the region, or the median of the intensities in the region.

**[025]** Finger touches are readily distinguishable from a fist, and an open hand. For example, a finger touch has relatively high intensity values concentrated over a small area, while a hand touch generally has lower intensity values spread over a larger area.

**[026]** For each frame, the system determines the number of regions. For each region, determine an area and location. The area is determined from an extent  $(x_{\text{low}}, x_{\text{high}}, y_{\text{low}}, y_{\text{high}})$  of the corresponding intensity values 104. This information also indicates where the surface was touched. A total signal intensity is also determined for each region. The total

intensity is the sum of the thresholded intensity values for the region. A time is also associated with each frame. Thus, each touched region is described by area, location, intensity, and time. The frame summary is stored in a hash table, using a time-stamp as a hash key. The frame summaries can be retrieved at a later time.

[027] The frame summaries are used to determine a trajectory of each region. The trajectory is a path along which the region moves. A speed of movement and a rate of change of speed (acceleration) along each trajectory can also be determined from the time-stamps. The trajectories are stored in another hash table.

[028] As shown in Figure 2A, the frame summaries 201 and trajectories 202 are used to classify gestures and determine operating modes 205. It should be understood that a large number of different unique gestures are possible. In a simple implementation, the basic gestures are no-touch 210, one finger 211, two fingers 212, multi-finger 213, one hand 214, and two hands 215. These basic gestures are used as the definitions of the start of an operating mode  $i$ , where  $i$  can have values 0 to 5 (210-215).

[029] For classification, it is assumed that the initial state is no touch, and the gesture is classified when the number of regions and the frame summaries remain relatively constant for a predetermined amount of time. That is, there are no trajectories. This takes care of the situation where not all fingers or hands reach the surface at exactly the same time to indicate a particular gesture. Only when the number of simultaneously touched regions remains the same for a predetermined amount of time is the gesture classified.

[030] After the system enters a particular mode  $i$  after gesture classification as shown in Figure 2A, the same gestures can be reused to perform other operations. As shown in Figure 2B, while in mode  $i$ , the frame summaries 201 and trajectories 202 are used to

continuously interpret 220 gestures as the fingers and hands are moving and touching across the surface. This interpretation is sensitive to the context of the mode. That is, depending on the current operating mode, the same gesture can generate either a mode change 225 or different mode operations 235. For example, a two-finger gesture in mode 2 can be interpreted as the desire to annotate a document, see Figure 5, while the same two-finger gesture in mode 3 can be interpreted as controlling the size of a selection box, as shown in Figure 8.

[031] It should be noted that the touch surface as described here enables a different type of feedback than typical prior art touch and pointing devices. In the prior art, the feedback is typically based on the  $x$  and  $y$  coordinates of a zero-dimensional point. The feedback is often displayed as a cursor, pointer, or cross. In contrast, the feedback according to the invention can be area based, and in addition pressure or signal intensity based. The feedback can be displayed as the actual area touched, or a bounding perimeter, e.g., circle or rectangle. The feedback also indicates that a particular gesture or operating mode is recognized.

[032] For example, as shown in Figure 3, the frame summary is used to determine a bounding perimeter 301 when the gesture is made with two fingers 111-112. In the case, where the perimeter is a rectangle, the bounding rectangle extends from the global  $x_{low}$ ,  $x_{high}$ ,  $y_{low}$ , and  $y_{high}$  of the intensity values. The center (C), height (H), and width (W) of the bounding box are also determined. Figure 4 shows a circle 401 for a four finger touch.

[033] As shown in Figures 5-9 for an example tabletop publishing application, the gestures are used to arrange and lay-out documents for incorporation into a magazine or a web page. The action performed can include annotating displayed documents, erasing the annotations, selecting, copying, arranging, and piling documents. The documents are



stored in a memory of a computer system, and are displayed onto the touch surface by a digital projector. For clarity of this description the documents are not shown. Again, it should be noted that the gestures here are but few examples of many possible gestures.

**[034]** In Figure 5, the gesture that is used to indicate a desire to annotate a displayed document is touching the document with any two fingers 501. Then, the gesture is continued by “writing” or “drawing” 502 with the other hand 503 using a finger or stylus. While writing, the other two fingers do not need remain on the document. The annotating stops when the finger or stylus 502 is lifted from the surface. During the writing, the display is updated to make it appear as if ink is flowing out of the end of the finger or stylus.

**[035]** As shown in Figure 6, portions of annotations can be “erased” by wiping the palm 601 back and forth 602 across on the surface. After, the initial classification of the gesture, any portion of the hand can be used to erase. For example, the palm of the hand can be lifted. A fingertip can be used to erase smaller portions. As visual feedback, a circle 603 is displayed to indicate to the user the extent of the erasing. While erasing, the underlying writing becomes increasingly transparent over time. This change can be on a function an amount of surface contact, speed of hand motion, or pressure. The less surface contact there is, the slower the change in transparency, and the less speed involved with the wiping motion, the longer it takes for material to disappear. The erasing terminates when all contact with the surface is removed.

**[036]** Figures 7-8 shows a cut-and-paste gesture that allows a user to copy all or part of a document to another document. This gesture is identified by touching a document 800 with three or more fingers 701. The system responds by displaying a rectangular selection box 801 sized according to the placement of the fingers. The sides of the selection box

are aligned with the sides of the document. It should be realized that the hand could obscure part of the display.

[037] Therefore, as shown in Figure 8, the user is allowed to move 802 the hand in any direction 705 away from the document 800 while continuing to touch the table. At the same time, the size of the bounding box can be changed by expanding or shrinking of the spread of the fingers. The selection box 801 always remains within the boundaries of the document and does not extend beyond it. Thus, the selection is bounded by the document itself. This enables the user to move 802 the fingers relative to the selection box.

[038] One can think of the fingers being in a control space that is associated with a virtual window 804 spatially related to the selection box 801. Although the selection box halts at an edge of the document 202, the virtual window 804 associated with the control space continues to move along with the fingers and is consequently repositioned. Thus, the user can control the selection box from a location remote from the displayed document. This solves the obstruction problem. Furthermore, the dimensions of the selection box continue to correspond to the positions of the fingers. This mode of operation is maintained even if the user uses only two fingers to manipulate the selection box. Fingers on both hands can also be used to move and size the selection box. Touching the surface with another finger or stylus 704 performs the copy. Lifting all fingers terminates the cut-and-paste.

[039] As shown in Figure 9, two hands 901 are placed apart on the touch surface to indicate a piling gesture. When the hands are initially are placed on the surface, a circle 902 is displayed to indicate the scope of the piling action. If the center of a document lies within the circle, the document is included in the pile. Selected documents are highlighted. Positioning the hands far apart makes the circle larger. Any displayed

documents within the circle hands are gathered into a 'pile' as the hands move 903  
towers each other. A visual mark, labeled 'pile', can be displayed on the piled documents.  
After documents have been placed in a pile, the documents in the pile can be 'dragged'  
and 'dropped' as a unit by moving both hands, or single documents can be selected by  
one finger. Moving the hands apart 904 spreads a pile of documents out. Again, a circle is  
displayed to show the extent of the spreading. This operation terminates when the hands  
are lifted from the touch surface.

**[040]** Although the invention has been described by way of examples of preferred  
embodiments, it is to be understood that various other adaptations and modifications may  
be made within the spirit and scope of the invention. Therefore, it is the object of the  
appended claims to cover all such variations and modifications as come within the true  
spirit and scope of the invention.